"For inclusion in the National Digital Catalog,
metadata shall include certain minimal data
describing the sites from which data and samples were collected."
(from FY 2009 Program Announcement)

The National Digital Catalog does not attempt to precisely describe every nuance of each sample-level data record but only seeks to include "minimal data" necessary to distinguish and describe the samples and encourage further exploration of similar items by identifying contacts and, where available, by linking to online resources capable of more fully describing the records.

Searches within the National Digital Catalog are thus limited to the information found in this minimal metadata. While the metadata is minimal the importance of each elements is significant and it is critical to understand each element and how it will be used in the National Digital Catalog.

Overview of Metadata Preparation

1. Understand the required and optional metadata accepted by the National Digital Catalog.
2. Map the National Digital Catalog metadata to the sample-level properties within your own existing data holdings.
3. Select a data upload format to use for upload to the National Digital Catalog.
4. Extract your existing holdings into the selected upload format.
5. Upload these extracted files into the National Digital Catalog.

Each of these steps are explained in detail below. Additional information is included in a separate document containing Frequently Asked Questions.

1 - Understanding the Metadata

The tables in this section list and discusses each of the thirteen (13) metadata elements accepted by the National Digital Catalog. The elements are separated by metadata which are required and must be present in the upload, and metadata which is optional but, when present, enhances the metadata record for users of the National Digital Catalog.

Required elements should always be included in the upload file with one exception. The CollectionID may be left blank and identified during the upload process. In this case, the CollectionID may be left blank.

Optional elements may be left out of the upload file.

Required Metadata

Required Element Name	Definition	#
collectionID[1]	A unique collection ID assigned by the National Digital Catalog to identify distinct collections. This field is required but may be left blank and assigned during the file loading process within the National Digital Catalog.	1
title	The human-readable title for the individual record that will be used in any listing or search result. It is best to keep this title short for display purposes but containing enough information to distinguish it from other records. **Sample:** Geologic Sample 160580 **Sample:** ID: NMDF52900064 TITLE: ISAACS BROS. LEAD-SILVER MINE **Sample:** Core Research Center, Cutting DD18216 **Sample:** Core sample from well: KNIK ARM ST 1	1
abstract	The human-readable description of the individual record used to help determine the nature of the underlying physical data resource. Due to the general nature of the catalog, a fair amount of information about the data resource may need to be captured into this one general element. **Sample:** Core Research Center, Cutting DD18210, from well operated by St. Michael Exploration, located in Weld County, CO, under lease 2-1 Grace State, with API number 0512310130. **Sample:** This is a geologic sample in one of the Bureau of Economic Geology's three Core Research Centers. API Number: 420513299400 Top Depth: 11744 Ft. Bottom Depth: 11767 Ft. sample_type_name: CORE CHIPS/CORE PLUGS sample_category_name: Core formation_name: Unknown formation_age_name: Unknown facility_name: Houston reservoir_name: BILOXI CREEK WILCOX operator_name: APACHE CORPORATION state_name: Texas county_name: Burleson	1
dataType[2]	A controlled vocabulary of data types. An item may include multiple dataTypes. These types include: • Auger Sample • Fluid Sample • Geochemical Sample • Hand Sample • Ice Core • Paleontological Sample • Rock Core • Rock Cuttings • Sediment Core • Sidewall Core • Thin Section	1-N

1 Collection IDs can be viewed by state at: http://my.usgs.gov/csc/nggdpp/state/** (where '**' is the 2-letter state abbreviation)
2 An extended list of types may be found at the end of the following document: http://datapreservation.usgs.gov/docs/NGGDPPMetadataProfile.pdf

Required Element Name	Definition	#
	• Type Stratigraphic Section	
supplementalInformation	This standard field will be used to provide specific information on how to access the physical data represented by the metadata record. This may be general for the entire collection (e.g., a URL to another Web site) or a specific reference to an online resource like an ordering system with a specific ID. **Sample:** Repository managed by the USGS Core Research Center, additional information can be found at http://geology.cr.usgs.gov/crc **Sample:** Web (this sample): http://inet1.beg.utexas.edu/crc2/geosample.aspx?ID=160580 Phone: 512-471-0402 (Austin CRC) Phone: 713-466-8346 (Houston CRC)	1
coordinates	Geographic longitude and latitude. Both values shall be contained in the same element and be listed in the order: longitude,latitude with values separated by a comma. **Sample:** -118.023423, 45.02312	1
datasetReferenceDate	A reference date indicating currency of the underlying data record. In many cases, this may be the date the metadata record was assembled for the National Catalog. Proper date formats are defined in ISO 8601 which include: • yyyy • yyyy-MM[3] • yyyyMMdd • yyyy-MM-dd	1

Optional Metadata

Optional Element Name	Definition	#
alternateTitle	Collection owners may elect to provide additional title identifiers for individual records for further identification or use by other Web service interfaces. The AlternateTitle field may include either textual titles or specific sample IDs used by the collection.	0 – N
alternateGeometry	The underlying collection resource may use an alternate method of identifying the location of the data. If so, this text field should be used to describe the authoritative source for geographic location and how the simple coordinates were derived. It is recommended that if the coordinates are not exact locations but calculated in some way that this alternateGeometry element be used to clarify that relationship and	1

3 YYYYMM is intentionally left out because it is apparently not allowed by ISO 8601.

Optional Element Name	Definition	#
	designate the source of the calculated points.	
onlineResource	One or more URL pointers to textual information about the specific record. Ideally, this should point to a web page that describes the specific sample in full detail, rather than a start page for a search mechanism.	0 – N
browseGraphic	One or more URL pointers to images representing the specific record.	0 – N
date	If a meaningful date within the geosciences domain can be attached to the record (e.g., a collection date), it can be supplied here. Either date may be to any degree of precision, or may be left blank to indicate uncertainty. **Sample:** 20090423 **Sample:** 1939-1945	0 - N
verticalExtent	If applicable, vertical extent information can be provided for the specific record. For example, vertical depth information can be very useful for rock core samples. Specification of extent can contain three elements: minimum value, maximum value, and unit of measure. For the purpose of the National Catalog, these elements will be collected as 2 or 3 values representing the UnitOfMeasure and MaximumValue with the possibile addition of MinimumValue (e.g., m,35.4,0 for a rock core measured at 35.4 total meters).	1

2 - Map Existing Metadata to National Digital Catalog Metadata

Having an understanding of the required and optional metadata elements accepted by the National Digital Catalog, you will now need to map the sample-level properties found in your own particular collection to these metadata elements.

In some cases this might be a one-to-one mapping where, for instance, the required “datasetReferenceDate” element might correspond to a '”DateCreated' property in your own collection database. In other cases, many properties from your own database may be used to populate a single metadata element. This could be the case for a required element such as “abstract” where it may be desirable to concatenate a number of fields to provide a richer description of the resource and more specific search results. In other cases, there may be a need to transform data from your database to a value suitable for the National Digital Catalog. For instance, if the location information were stored in township/range/section, it would be necessary to convert these values to a single geographic latitude and longitude point to populate the required coordinates metadata element.

Included on the following page is a blank table which may be printed and used as a worksheet for the purpose of mapping your existing data to the required metadata fields.

National Catalog Element	Mapped Property or Properties	Transformaton?
Required Elements (only dataType is repeatable and is encoded as a comma delimited list of values in a single element or column)		
collectionID		
title		
abstract		
dataType		
supplementalInformation		
coordinates		
datasetReferenceDate		
Optional Elements (repeatable elements shown twice)		
alternateTitle		
alternateTitle		
alternateGeometry		
onlineResource		
onlineResource		
browseGraphic		
browseGraphic		
date		
date		
VerticalExtent - unit of measure - maximum value - minimum value		

3 - Select an Upload Format

The National Digital Catalog supports two primary formats for loading data. The simplest format is commonly referred to as a Comma Separated Value or CSV format. This format is fairly easy to create and many software packages use this format to exchange tabular data. The more complex format accepted by the National Digital Catalog is a custom XML format that provides better handling for repeatable elements than possible with the more simple CSV format. The XML also eliminates some of the parsing issues with the CSV format. Each of these formats will be discussed in more detail in the sections below.

CSV File Format

A CSV file format represents tabular data in such a way that each line in the file corresponds to a single record, and the properties associated with each record are delimited on the same line by a comma character. The comma, in this case is a special character that separates data fields. This is useful for many applications but causes problems when the data field that is being delimited itself includes a comma. It then becomes necessary to enclose the entire field in double quotes so that the embedded comma is not seen as a field delimiter but simply as part of the field's information. This works well unless the information in the field also uses a double quote as well as a comma to indicate (for the sake of example) inches in a measurement. In general, the comma separated value actually doesn't work well for many types of data. Fortunately, the National Digital Catalog allows use of another less common special character to delimit record values.

It is recommended that the “CSV” file actually use what is called the “record delimiter” or “pipe” character (|) to delimit fields rather than a comma. This character is seldom used in normal text and allows fields to include the more common comma and quote characters without risking confusion with the field separator. For instance, a single record with four fields (more are required but trimmed in this example to avoid line wrapping) might look like this using the record delimiter character:

```
1341234|This is a Title|m,35.4,0|Please note, the existence of a comma and "quote" characters.
```

The above line could not be represented accurately using a comma delimited approach but is fairly straightforward using the chosen record delimiter character.

It is also necessary to include in the “CSV” formatted file a single record on the first line that indicates the metadata element names corresponding to the subsequent data. Rather than impose a strict order of elements requirement the loader simply requires that the first line of the CSV include the element names in a similar delimited fashion. Applying this to the above example, the first two lines of the CSV file would look something like the following:

```
COLLECTIONID|TITLE|VERTICALEXTENT|ABSTRACT
1341234|This is a Title|m,35.4,0|Please note, the existence of commas and quote characters.
```

XML File Format

The XML file format is more complex than the CSV, can better handle repeating values for the same record and is normally more reliable for handling various data inputs. It is suitable for those who are familiar with the XML format and have tools available to work directly with this format. It is not recommended for users who have no previous experience with XML.

When using the XML file format, new elements are added to the required and optional metadata elements listed above to handle repeatable elements. For instance, the alternateTitle element has one-to-many "title" child elements to handle multiple alternateTitle elements.

An online XML template can be found at: http://datapreservation.usgs.gov/docs/collectionMetadataExample.xml

A very basic two-record sample is also available at: http://datapreservation.usgs.gov/docs/NGGDPPSampleMetadata1.xml

And an XML schema is provided for basic XML validation: http://datapreservation.usgs.gov/docs/NGGDPPMetadata.xsd

These examples should be sufficient to understand the structure of the XML file the National Digital Catalog expects to receive during a load.

4 - Extract Sample-Level Data into Upload Format

The extraction process simply transforms the records in your collection into National Digital Catalog metadata elements and stores this in the selected file format (CSV or XML). Obviously, a repeatable process is desirable as this process will be needed for subsequent updates.

Most databases have the ability to write out simple "CSV" formatted files through built-in utilities, or customized queries. Excel may also be used to write out a "CSV" formatted files but the options for character encoding and delimiters are somewhat limited. Excel uses the system settings to determine character encoding and delimiters. On a Windows platform these can be set by accessing the Control Panel --> Regional and Language Options.

Some databases may also have the capability to extract information in an XML format. This would not be the formatted expected by the National Digital Catalog but could be transformed using an additional XSL stylesheet into the correct format.

It is not within the scope of this document to explore all of these options. It is the responsibility of the data steward to provide this extraction using the best means at their disposal. We will add extraction helps as we become aware of them for different systems.

5 – Upload to National Digital Catalog

Once a correctly formatted extract file is available, the file will need to be loaded into the National Digital Catalog through the provided web interface. This interface requires the data loader to login to the myUSGS system and be given the necessary permission to edit and add data to the collection(s) indicated. Here are some guidelines:

- First, you will need a login account on the myUSGS collaboration system and the account must have specific permissions established to allow access and upload to the National Digital Catalog. This account may be requested by sending email to myusgs@usgs.gov with a request similar to the following:

I am a data steward for a geophysical/geological collection which is part of the NGGDPP. I am requesting a myUSGS user account with NGGDPP_Author role to upload sample metadata to the National Digital Catalog. My contact information is:

Agency:
First Name:
Middle Initial:
Last Name:
Work Email Address:
Work Phone Number:

- After receiving a myUSGS user account, you should be able to follow the upload steps provided at the following site:

  http://my.usgs.gov/csc/nggdpp/upload

Contact Information

Any questions or concerns can be addressed to the following contacts:

Richard Brown
USGS - Central Region Geospatial Information Office
1400 Independence Drive, Rolla, MO 65401
573-308-3525
reb@usgs.gov

Frances Wahl Pierce
National Geological and Geophysical Data Preservation Program
U.S. Geological Survey, 912 National Center, Reston, VA 20192
703-648-6636
fpierce@usgs.gov